QUALITY CONTROL AND CHECKING OF ELECTRONICALLY SUBMITTED PAPERS TO ACCELERATOR CONFERENCES *

William W. MacKay, BNL, Upton NY

Abstract

Recent conferences and workshops have been producing proceedings in paper and electronic forms in both CDROM and Web formats. In order to have a readable paper with some uniformity in a given publication, certain rules must be followed by authors to ensure a usable final product. This paper discusses the quality control process.

1 BASIC REQUIREMENTS

Each paper should be readable in both the paper copy and on the computer screen. The most important thing is to convey the information which the author intended. The electronic copy should be printable with the same or better quality as the paper proceedings. Even though some color is allowed for the electronic copy, no information should be lost when the paper is printed in gray scales to a black and white printer.

The electronic files should be as small as possible, since people will sometimes download the files across slow phone lines and may also have computers with limited memory and CPU speeds. It is particularly irksome to wait many minutes to download and display a huge PDF file only to discover that you are not really interested in the paper. We have found that some large files may take several minutes just to print, if they print at all.

The electronic copy should display on as many computer platforms as possible. To this end, the Adobe portable document format (PDF) has been adopted for the final electronic version. Since there are a number of ways to generate PDF files with varying levels of quality, at present we ask for PostScript files rather than PDF files. The electronic processing team then distills the PostScript files into PDF files with Acrobat Distiller using appropriate settings. Even so there can be problems with the PostScript file, so we must also ask the authors for the electronic source files (LaTeX, MSWord, figures, $\cdots$).

2 STEPS IN PROCESSING

Before discussing the quality assurance in detail it is necessary to outline the steps in producing the proceedings.

1. Each PostScript file is distilled via Adobe Acrobat Distiller into a PDF file.

2. The PDF file is checked with Adobe Acrobat Exchange for QA with some minor massaging including cropping of margins.

3. The Adobe Acrobat Reader is then used to convert the PDF files back into PostScript files.

4. A Perl script adds headers and footers with the conference title and page numbers to these PostScript files.

5. The PostScript files with page numbers are redistilled into PDF files.

6. Thumb nails are added to the PDF files

7. A searchable index is built from the PDF files.

8. The table of contents and author index are generated from the database.

9. Other boiler plate for the proceedings is generated.

10. The number of volumes and page breaks are determined.

11. All the papers are printed and assembled into proof copies of the printed volumes.

12. For PAC'99, we sent a hard copy to the publisher to print the paper version of the proceedings, although many publishers will also accept files in an electronic format. [1]

13. The printer makes a galley proof of the volumes and sends it back for checking. Corrections to the galley proofs are returned to the publisher.

14. The proceedings are printed and mailed.

15. A web site is built for the electronic version of the proceedings. Authors should be requested to check their papers for errors on the web.

16. The boiler plate is converted to HTML.

17. Web versions of the table of contents and author index are generated from the database.

18. The web pages of links for keywords are also generated from the database.

19. The web site is copied to a CDROM and sent to the publisher for duplication and mailing.

---

[1] Since we had already assembled and paginated the complete text, we thought it would be simpler to send a hard copy. Then we only needed to replace a few pages after reviewing the galley proofs.

3 STEPS OF CHECKING

The first check should be made by the author. It is best if the conference can set up an automated Distiller for authors to test the distillation of papers from PostScript to PDF. The author should then retrieve the PDF file and check it for readability, printability, and format including margins. After correcting problems the author should then submit the paper as a PostScript file along with the source files (Word, LaTeX, figures, etc.) and a description of software and platform used to generate the paper.

Before the papers are processed the Distiller needs to be set up with correct parameters on **all** computers which will be used by the processing team. At PAC'99 we had some problems with this and had to reprocess some papers. If individual accounts are given to the processing team, we need to ensure that the distiller profile for each account is set up identically on each computer, since people will not always sit at the same computer.

3.1 First pass

1. Distill the PostScript file with correct settings. This pass should embed the complete fonts and not subsets, since there will be a second pass through the distiller. We have seen an occasional paper which looks fine after this step, but after the page numbers have been added, some of the symbols (such as a vector sign) in equations get substituted by letters from the Courier font.

2. Check the file size of the PDF file. Files much larger than a megabyte can cause problems. (For PAC'99 the average PDF file size was 171 kbytes.)

3. View the paper on the screen. Note how long the paper takes to display. If some figures display too slowly, then the author may have to rework the figures and resubmit the paper.

4. Check the title and authors against the traveler; sometimes files end up in the wrong directories or with incorrect names. Minor modifications to the title should be OK, so long as the subject is the same. The list of authors for a paper is frequently modified, but the submitting author is generally the same.

5. Check the page limit; if it is too long for the type of paper, then have the author shorten and resubmit the paper.

6. Check that the correct Type 1 fonts have been used. There may be a few Type 3 fonts, particularly in figures, or if BlueSky fonts have not been installed with LaTeX. In the past if the Type 3 fonts appear in only a few places, then they have been accepted; otherwise the paper will have to be resubmitted or reprocessed by an expert from the source files.

7. Check the font list for font substitutions (see §4.2).

8. Check that the paper has been formatted according to the instructions.

9. Check the margins. In many cases it may be simpler to shift things a little with PitStop than to ask the author to resubmit.

10. Print the file, then check the margins with a template, and compare with the author's original.

If it passes all these tests, then the paper is ready to be given to the QA team.

3.2 Second Pass (QA)

The QA team should quickly check the same things as mentioned in the above list. The QA team should should be made up of only a few people who will review the formatting of the papers in a consistent manner. They provide another set of eyes to view the paper to see if something has been overlooked. It is useful to have people who are used to reading mathematical equations scan over each paper to see if something looks amiss. If the QA team finds something wrong, the paper should be sent back to the main processing team, rather than wasting time in QA. Just as it is important to educate authors, it is also important to educate the paper processors, so they do not keep repeating the same mistakes.

3.3 After Pagination

When the printed copy has been assembled with page numbers, each page should be checked for flaws and that the pages are numbered consecutively. The table of contents and author index should also be checked against the page numbers.

Every page of galley proofs should be scanned, and corrections sent back to the publisher. For a large publication, there will always be yet another error which you do not find.

The web version needs to have the links checked for the table of contents, author index, and keywords. It is fastest to have a large group of people check the links of the papers by giving each person a different set of links and papers to check.

After producing a good online version, then a prototype CDROM may be made. Several copies should be made and tested on as many platforms as are available. For PAC'99, we tested the CDROM on a PC with Windows, PC with Linux, Macintosh, Sun, and HP Unix workstations. Several passes may be necessary to generate a working CDROM.

After sending the CDROM off to the publisher, a sample should be checked before mailing.

4 FORMAT

The paper should be formatted as explained in the JACoW instructions for preparation of papers for accelerator conferences. The JACoW templates are set up to use the correct format; although, sneaky authors may override or try

to change the format. The editing team must decide how flexible they will be on certain things. For example, if the template requires titles in upper case only, you may find a lot of time is spent changing titles to upper case when half of the papers have mixed case titles. (PAC'99 actually tried to enforce this, which was probably a mistake; although, it may make some authors pay more attention to the instructions for future conferences.)

The important things to watch are that the paper is readable, printable, within the page limits and margins, and that the general appearance is at least close to the required style.

4.1 Page Limits and Margins

So long as there are printed proceedings, there will probably be page limits for papers at some conferences. It is particularly important to enforce the page limit for large conferences. (PAC'99 had 1223 papers to publish.)

Margins are a bit tricky, since two different paper sizes are used: A4 (21.0 cm × 29.7 cm) and US Letter (8.5 in × 11 in). A method has been devised to convert the PDF file to print on both sizes of paper with reasonable margins. The PDF file should be cropped down to a smaller size than either paper size and with equal left and right margins and equal top and bottom margins. PitStop has sets of guidelines which may be defined and loaded to highlight the desired boundary of the text area. Typically, for papers generated for US Letter size, 6 mm should be cropped from the right hand side. Papers generated for A4 size, should have 18 mm cropped from the top. Not all papers will conform to these two cases. (I have found some LaTeX installations at BNL set up with dvips set up for A4 size paper, even though we only have Letter size printers.) After cropping, the text of the paper should be centered in a rectangle of 210 mm (horizontal) and 279 mm (vertical) as shown in the Acrobat Exchange window. In some cases it may have been necessary to crop the paper more tightly, but the text area should still be centered in the remaining white rectangle. Even if the paper has been cropped to a smaller rectangle, the correct margins will be restored when the PDF file is converted to a PostScript file to add the page numbers.

Sometimes text or figures extend past the margins. One of the commonest problems is that the MSWord table with outlines frequently sticks out into the margin; I consider this to be a bug in Word. Some figures may be moved with PitStop; however, if the figure has a lot of vectors, PitStop may not be able to handle it. Some lines which extend past the margins may be truncated by cropping the page down to the edge of the text area.

4.2 Fonts

Type 1 fonts should be used. Bit-mapped fonts, such as the Adobe Type 3 fonts do not work well at resolutions which differ from the intended size. LaTeX installations should be upgraded to include the BlueSky Type 1 versions of the Computer Modern (CM) and AMS fonts. If a PostScript file that contains Type 3 fonts was generated from LaTeX,

then a simple fix which usually works is to run the source files through LaTeX on a computer with the BlueSky fonts installed.

For plain text, the Adobe Times, Symbol, and ZapfDingbats font sets should be used for the paper. In the past we have made the mistake of saying these are the only allowed fonts; however, when it comes to equations a lot of symbols would then be missing. (In fact, the Greek letters in the Symbol font are not slanted as in the usual case of high quality mathematical publications.) The LaTeX template invokes the Computer Modern (CM) fonts rather than the Symbol or ZapfDingbats fonts in math mode.

Additionally a fixed-size typewriter font is desirable for examples of computer code. A fixed-width font with serifs is recommended, so that the code lines up as in a simple editor. Without serifs, a capital i, a lower case el and the number one may be confused. The Adobe Courier font is an acceptable font; although it is a little larger than the Times set. With LaTeX the CMTT font is better, since it is slightly smaller.

With fonts in figures, the author may not have as much control, particularly if the figure is already a bit map. Fonts for figures should be at least large enough to be easily read. It is recommended that the size of fonts in figures as printed in the paper (after rescaling the figure) is no smaller than about 5 or 6 points.

Font substitutions occur when some of the fonts have not been included in the PostScript file. Frequently this is seen in figures, or files generated by Textures on a Macintosh or by Asian versions of MSWord. The usual symptom is that the default font (usually Courier) appears in some odd places. Additional symptoms are scrambled fonts with some overprinting and incorrect characters in equations. Missing fonts get substituted by the default font (usually Courier, but it depends on the Distiller preferences). When the metric gets fouled up, some characters overprint and may even appear in the wrong order.

Depending on the severity of the problem, several solutions are possible:

1. Get the author to fix it and resubmit.

2. Regenerate the PostScript file from the original source files, but using correct fonts. This usually works quite well with LaTeX, but MSWord is more likely to have text flow over onto an extra page. The other problem which occurs is that the fonts which were missing are also not in the set of fonts available to the paper processor.

3. If only a few characters have been scrambled, then it is frequently simpler to edit the file with PitStop; the font for an individual character may be changed, and individual characters and words may be dragged around.

Whichever method is used to fix the problem, the final output should be checked carefully against the author's original paper copy.

4.3 Titles and Headings

The titles and section headings should follow the format given in the instructions for preparation of papers. It may improve the appearance of the paper to adjust the vertical location of some headings with PitStop.

One of the deficiencies in the LaTeX template is that space allocated for the author list may need to be tweaked. With multiple lines of authors, there may not be enough margin between the author list and the beginning of the text, but there is usually a bit of extra space between the title and author list.

Misspelled words in titles and headings frequently stand out. They can be corrected in Acrobat Exchange.

Sometimes authors do not use the correct fonts and alignment in the headings. These are frequently fixable with PitStop.

4.4 Figures

Of the time spent processing papers, figures tend to cause the most grief. This is because there are so many kinds of figures and ways to produce them. Various canned software is used, and some of the programs generate bad encapsulated PostScript files.

Large file sizes are due to figures. High density bit maps can be quite large, but in many cases they may distill to a decent size. Poincaré and other plots with a large number of vectors may run to several megabytes in PostScript and will not compress well with the Distiller. Such figures may be converted to bit maps which will then compress to smaller sizes, although some of the resolution may be lost, particularly for labels.

In the past a compression setting of around 200 dpi was recommended for compressing figures. This works well for photographs and some other figures, but some bit maps require a higher density to print well. (Martin Comyn suggested using a density of 600 dpi.) Frequently a bit mapped figure which prints well may be grainy when displayed on the screen. If the full information of a figure displays acceptably when the window is zoomed in, then it is deemed to be all right.

4.5 Page Numbers and Headings

For PAC'99, we found that the Perl script worked very well with only a couple of failures. In one case a figure with improper encapsulated PostScript rescaled to cover up some of the text. In the few cases, we were able to add the page numbers by hand with PitStop.

5 CONCLUSION

Processing and checking papers for a large conference takes a considerable amount of time. In the old method of submitting camera ready papers, most of the quality assurance was done by the author. With electronic publication the author still can do quite a bit of the QA if a automated test site is set up, but due to the complexities of electronic files there is still quite a lot of work to be done after a paper is submitted. Things are improving, and with a common template and instructions for various conferences, a lot of the current problems should be reduced. The problems with fonts are fairly well understood and should diminish as people become aware and upgrade their systems to use the correct fonts. The problem with figures will probably remain for a while, but hopefully more people will make an effort in the future to clean up their figures before submitting papers.